ISRAEL JOURNAL OF MATHEMATICS, Vol. 67, No. 2, 1989

ON THE CLASSIFICATION OF ONE-SIDED MARKOV CHAINS

BY
R. COWEN
Mathematics Department, University of Nairobi, P.O. Box 30197, Nairobi, Kenya

ABSTRACT
In this paper we generalise a theorem of R. F. Williams about the topological classification of one-sided finite state stationary Markov chains. Specifically, we give a classification up to block-isomorphism, i.e. a topological conjugacy between one-sided Markov chains which preserves the Markov measures.

§0. Introduction

The theory of topological Markov chains (or subshifts of finite type) plays an important role in many branches of ergodic theory and dynamical systems. We will be concerned with the measure-theoretic classification of the one-sided Markov chains.

In [3] Williams introduced a complete invariant for topological conjugacy between one-sided topological Markov chains and he introduced (different) complete invariants for topological conjugacy between two-sided topological Markov chains. Unlike the situation for the two-sided case, there is a finite procedure for determining whether two one-sided topological Markov chains are topologically conjugate.

The topological classification of two-sided topological Markov chains was extended by Parry and Williams (cf. [2]) to give a classification of two-sided Markov chains up to block-isomorphism. That is, they gave complete invariants for two-sided Markov chains to be topologically conjugate by a conjugacy which preserves the Markov measures. The invariant introduced in [2] was further refined by Parry and Tuncel in [1] using matrices whose entries

Received November 15, 1988 and in revised form January 19, 1989

are non-negative integral combinations of exponential functions. By working over the same semi-ring we will extend the topological classification of one-sided topological Markov chains to give a classification up to block-isomorphism. That is, we will give complete invariants for one-sided Markov chains to be topologically conjugate by a conjugacy which preserves the Markov measures. With this generalisation of Williams's theorem there still remains a finite procedure for determining whether one-sided Markov chains are block-isomorphic.

I wish to thank Professor William Parry for his guidance and encouragement throughout the course of this work

§1. Markov chains

Let S be a $n \times n$ irreducible 0–1 matrix. Give $\{1, \dots, n\}$ the discrete topology and $\Sigma^+ = \Pi_0^\infty \{1, \dots, n\}$ the product topology. Consider the subspace $\Sigma_S^+ \subset \Sigma^+$ defined by

$$\Sigma_S^+ = \{x \in \Sigma^+ : S(x_i, x_{i+1}) = 1 \text{ for all } i \geqq 0\}.$$

The shift σ_S^+ is defined on Σ_S^+ by $(\sigma_S^+ x)_i = x_{i+1}$ for $x = (x_i)$. σ_S^+ is a bounded-to-one continuous surjection and (Σ_S^+, σ_S^+) is called a *one-sided topological Markov chain* (or *subshift of finite type*).

Given two one-sided topological Markov chains (Σ_S^+, σ_S^+) and (Σ_T^+, σ_T^+) we say that they are *topological conjugate* if there exists a homeomorphism ϕ of Σ_S^+ onto Σ_T^+ such that $\phi\sigma_S^+ = \sigma_T^+ \phi$.

Let P be a stochastic matrix and denote the matrix obtained from P by raising every non-zero entry to the power $t, t \in R$ by P^t. Let p denote the unique probability vector such that $pP = p$. From the stochastic matrix P we can define a unique (Markov) probability measure m_P on Σ_{P^0} where m_P is σ_{P^0}-invariant. This is defined on the Borel subsets of Σ_{P^0} and assigns $p(i_0)P(i_0, i_1) \cdots P(i_{n-1}, i_n)$ to the cylinder

$$| i_0, \dots, i_n |^m = \{x \in \Sigma_{P^0} : x_m = i_0, \dots, x_{m+n} = i_n\}.$$

§2. Block-isomorphism

In this paper we will give necessary and sufficient conditions for two one-sided Markov chains to be topologically conjugate by a homeomorphism that preserves the Markov measures. The proof of this theorem (Theorem 2) will be given in §3 and will closely follow the proof of Theorem 11 given by Williams in [3].

A rectangular 0–1 matrix is called a *division matrix* if its rows are non-trivial and each column contains exactly one non-zero entry. A square 0–1 matrix is called a *permutation matrix* if its rows and columns contain exactly one non-zero entry.

For matrices S and T we say that *S is a reduction* of T and write $S < T$ if there exist non-negative integer matrices D and R, where D is a division matrix, such that

$$S = DR \quad \text{and} \quad T = RD.$$

Given a matrix S we say that *S_0 is a total reduction of S* provided:

(i) $S_0 < A_1 < \cdots < A_r = S$ for some sequence of matrices,

(ii) S_0 has no repeated column.

Williams proved the following classification theorem:

THEOREM 1 [3] (Williams). *Every square matrix S over Z^+ has a total reduction S_0. (Σ_S^+, σ_S) and (Σ_T^+, σ_T) are topologically conjugate if and only if their total reductions S_0, T_0 are conjugate by a permutation (i.e. $S_0 = PT_0P^{-1}$ for a permutation matrix P).*

We will extend this classification up to *block-isomorphism*, i.e. a topological conjugacy that preserves Markov measures. Let $Z^+(\exp)$ denote the semi-ring of positive integral combinations of exponential functions. Let $P(t)$ and $Q(t)$ be square matrices with no trivial rows or columns whose entries are in $Z^+(\exp)$. Also suppose that $P(1)$ and $Q(1)$ are stochastic. We say that *$Q(t)$ is a reduction of $P(t)$* and write $Q(t) < P(t)$ if there exist rectangular matrices $R(t)$ and $D(t)$ whose entries are in $Z^+(\exp)$ such that $R(1)$ and $D(1)$ are stochastic, $D(0)$ is a division matrix and

$$P(t) = R(t)D(t), \qquad Q(t) = D(t)R(t).$$

If there exist matrices $P_0(t), \ldots, P_n(t)$ such that $P_0(t) = P(t)$, $P_n(t) = Q(t)$ and for each $1 \leqq i \leqq n-1$ either $P_i(t) < P_{i+1}(t)$ or $P_{i+1}(t) < P_i(t)$, we say that *$P(t)$ and $Q(t)$ are related*. Given a matrix $P(t)$, a *total reduction* is a matrix $P_0(t)$ satisfying:

(i) $P_0(t) < P_1(t) < \cdots < P_n(t) = P(t)$ for some sequence of matrices,

(ii) $P_0(t)$ has no column which is some exponential c^t times another column.

We will prove:

THEOREM 2. *Every stochastic matrix P has a total reduction $P_0(t)$, $(\Sigma_{P^0}^+, \sigma_{P^0}, m_P)$ and $(\Sigma_{Q^0}^+, \sigma_{Q^0}, m_Q)$ are block-isomorphic if and only if the total reductions $P_0(t)$ and $Q_0(t)$ are conjugate by a permutation ($P_0(t) = S^{-1}Q_0(t)S$).*

§3. Proof of Theorem 2

To prove Theorem 2 we require a series of lemmas.

Given an $n \times n$ matrix S the state partition α_S of Σ_S^+ is the partition into sets $|i|^0$ for $1 \leqq i \leqq n$ where

$$|i|^0 = \{x \in \Sigma_S^+ : x_0 = i\}.$$

Suppose that α and η are partitions of Σ_S^+, then we write $\alpha \leqq \eta$ if every element of the partition α is a union of elements of η.

For $n \geqq 0$ let

$$\alpha \vee \sigma_S^{-1}\alpha \vee \cdots \vee \sigma_S^{-n}\alpha = \{A_0 \cap \cdots \cap A_n : A_i \in \sigma_S^{-i}\alpha, 0 \leqq i \leqq n\}$$

and denote this partition by α^n. We shall need the following lemma:

LEMMA 1 [2]. *Suppose that P is an irreducible stochastic matrix, let β and η be partitions of $(\Sigma_{P^0}^+, \sigma_{P^0}, m_P)$ into closed-open sets and suppose $\alpha \leqq \eta \leqq \alpha^1$. Define two stochastic matrices indexed by $\alpha \times \eta$:*

$$|\alpha, \eta|(K, E) = \frac{m_P(K \cap E)}{m_P(K)} \quad \textit{and} \quad |\alpha, \eta|_{\sigma_{P^0}}(K, E) = \frac{m_P(K \cap \sigma_{P^0}^{-1} E)}{m_P(K)}$$

for $(K, E) \in \alpha \times \eta$; then

$$|\alpha, \eta|\,|\eta, \alpha|_{\sigma_{P^0}} = |\alpha, \alpha|_{\sigma_{P^0}} \quad \textit{and} \quad |\eta, \alpha|_{\sigma_{P^0}}|\alpha, \eta| = |\eta, \eta|_{\sigma_{P^0}}.$$

Note that $|\alpha, \eta|$ is a division matrix and that the products $|\alpha, \eta|^t|\eta, \alpha|_{\sigma_{P^0}}^t$ and $|\eta, \alpha|_{\sigma_{P^0}}^t|\alpha, \eta|^t$ are 0–1 matrices when $t = 0$.

LEMMA 2. *If P and Q are stochastic matrices and $\phi : \Sigma_{P^0}^+ \to \Sigma_{Q^0}^+$ is a block-isomorphism, then P^t and Q^t are related.*

PROOF. Let $\eta = \phi^{-1}\alpha_{Q^0}$ and choose n such that $\eta \leqq \alpha_{P^0}^n$ and $\alpha_{P^0} \leqq \eta^n$. Consider the following sequence of partitions:

$$\alpha_{P^0} \vee \eta^{n-1} \leqq \eta^n \leqq (\alpha_{P^0} \vee \eta^{n-1})^1,$$

$$\alpha_{P^0} \vee \eta^{n-2} \leqq \alpha_{P^0} \vee \eta^{n-1} \leqq (\alpha_{P^0} \vee \eta^{n-2})^1,$$

$$\cdots \qquad \cdots \qquad \cdots$$

$$\alpha_{P^0} \vee \eta \leqq \alpha_{P^0} \vee \eta^1 \leqq (\alpha_{P^0} \vee \eta)^1.$$

By raising each of the matrices defined in Lemma 1 to the power t, we have that $|\eta^n, \eta^n|_{\sigma_{P^0}}^t$ and $|\alpha_{P^0} \vee \eta, \alpha_{P^0} \vee \eta|_{\sigma_{P^0}}^t$ are related. Similarly $|\alpha_{P^0}^n, \alpha_{P^0}^n|_{\sigma_{P^0}}^t$ and $|\alpha_{P^0} \vee \eta, \alpha_{P^0} \vee \eta|_{\sigma_{P^0}}^t$ are related. Now $|\alpha_{P^0}, \alpha_{P^0}|_{\sigma_{P^0}}^t = P^t$ and $|\eta, \eta|_{\sigma_{P^0}}^t = Q^t$. These matrices are clearly related to $|\alpha^n_{P^0}, \alpha^n_{P^0}|^t_{\sigma_{P^0}}$ and $|\eta^n, \eta^n|^t_{\sigma_{P^0}}$, respectively. Hence P^t and Q^t are related.

We now show how total reductions can always be found.

LEMMA 3. *Let $P(t)$ be a square matrix with no trivial rows or columns such that $P(1)$ is stochastic and whose entries are in $Z^+(\exp)$. Then we can find a total reduction $P_0(t)$ of $P(t)$.*

PROOF. Let $P(t)$ be a $n \times n$ matrix and suppose column $j = c^t \times$ column i. Let the integer k vary over the set $\{1, \ldots, i, \ldots, j-1, j+1, \ldots, n\}$. Define a $n \times (n-1)$ matrix $R(t)$ where column k of $R(t)$ equals column k of $P(t)$ if $k \neq i$. When $k = i$ let column k of $R(t)$ equal $(1+c) \times$ column i. Now let D' be the $(n-1) \times n$ division matrix that partitions the standard row vectors that generate $\mathbf{Z}^n$, $\{y_1, \ldots, y_n\}$ into $n-1$ sets $\{U_1, \ldots, U_{n-1}\}$ where $U_k = \{y_k\}$ for $k \neq i$ and $U_k = \{y_i, y_j\}$ for $k = i$. ($D_{lm} = 1 \Leftrightarrow y_m \in U_l$.) We now construct $D(t)$ by altering the unique non-zero entry of column i in D' to $1/(1+c)^t$ and changing the unique non-zero entry of column j of D' to $c^t/(1+c)^t$. Then $D(1)$ and $R(1)$ are stochastic matrices with $P(t) = R(t)D(t)$. We now repeat this procedure if necessary on $D(t)R(t)$. Since the size of our matrices are being reduced every time this procedure is followed, we will eventually obtain a total reduction of $P(t)$.

Given two reductions of some matrix we can always find a matrix which is a reduction of both of them:

LEMMA 4. *Suppose $B(t)$ and $C(t)$ are reductions of $P(t)$, then there exists $A(t)$ which is a reduction of both $B(t)$ and $C(t)$.*

PROOF. Let $P(t)$ be $n \times n$, $B(t)$ $m \times m$ and $C(t)$ $r \times r$. Let q be the smallest integer such that the columns of $P(t)$ can be partitioned into two sets W_1 and W_2 of q and $n-q$ columns respectively, where each column of W_2 is some exponential (c^t, for $c > 0$) times one of the columns in W_1. Express $P(t)$ as a product $R_1(t)D_1(t)$ where $R_1(t)$ is $n \times q$ and $D_1(t)$ is $q \times n$ by the method used in Lemma 4 and put $A(t) = D_1(t)R_1(t)$.

Since $B(t)$ is a reduction of $P(t)$ there are matrices $D_2(t)$ and $R_2(t)$ such that

$$P(t) = R_2(t)D_2(t) \quad \text{and} \quad B(t) = D_2(t)R_2(t).$$

We claim that there exists a $q \times m$ matrix $D_3(t)$ such that $D_3(0)$ is division, $D_3(1)$ is stochastic and $D_1(t) = D_3(t)D_2(t)$. Let the standard row vectors which generate $\mathbf{Z}^n$, $\mathbf{Z}^q$ and $\mathbf{Z}^m$ be $\{x_1, \ldots, x_n\}$, $\{y_1, \ldots, y_q\}$ and $\{z_1, \ldots, z_m\}$ respectively. The division matrix $D_1(0)$ gives a partition $\{U_1, \ldots, U_q\}$ of

$\{x_1, \ldots, x_n\}$ and similarly $D_2(0)$ gives a partition $\{V_1, \ldots, V_m\}$ of $\{x_1, \ldots, x_n\}$. Now for each $1 \leqq j \leqq m$ the columns of $P(t)$ corresponding to all the x_k's in V_j are an exponential times each other. But $\{U_1, \ldots, U_q\}$ is the smallest partitioning of the x_k's into sets whose corresponding columns are an exponential times each other. Thus $\{V_1, \ldots, V_m\}$ refines $\{U_1, \ldots, U_q\}$. Let D_3 be the division matrix that partitions $\{z_1, \ldots, z_m\}$ into sets $\{Y_1, \ldots, Y_q\}$ where $Z_j \in Y_i$ if $V_j \subset U_i$; then $D_1(0) = D_3 D_2(0)$. If $D_3(i,j) = 1$ and $x_k \in V_j \subset U_i$, then $D_1(t)(i,k) \neq 0$, $D_2(t)(j,k) \neq 0$ and we can define $D_3(t)$ by

$$D_3(t)(i,j) = \frac{D_1(t)(i,k)}{D_2(t)(j,k)} .$$

We must check that this definition is unambiguous so suppose $x_k, x_l \in V_j \subset U_i$. Choose s such that $P(t)(s,k) \neq 0$; as $P(t)(s,k) = R_1(t)(s,i)D_1(t)(i,k)$ we have that $R_1(t)(s,i) \neq 0$. Now

$$\begin{aligned} D_1(t)(i,k)D_2(t)(j,l) &= \frac{P(t)(s,k)D_2(t)(j,l)}{R_1(t)(s,i)} \\ &= \frac{R_3(t)(s,j)D_2(t)(j,k)D_2(t)(j,l)}{R_1(t)(s,l)} \\ &= \frac{P(t)(s,l)D_2(t)(j,k)}{R_1(t)(s,i)} \\ &= D_1(t)(i,l)D_2(t)(j,k). \end{aligned}$$

Hence

$$\frac{D_1(t)(i,k)}{D_2(t)(j,k)} = \frac{D_1(t)(i,l)}{D_2(t)(j,l)}$$

and $D_3(t)$ is defined unambiguously.

Clearly $D_1(t) = D_3(t)D_2(t)$ and $D_3(1)$ is stochastic since $D_1(t) = D_3(t)D_2(t)$.

Define $R_3(t) = D_2(t)R_1(t)$ and $A(t) = D_3(t)R_3(t)$, then

$$\begin{aligned} D_2(t)R_2(t)D_2(t) &= D_2(t)R_1(t)D_1(t) \\ &= D_2(t)R_1(t)D_3(t)D_2(t). \end{aligned}$$

Since each column of $D_2(t)$ contains only one non-zero entry we conclude that $D_2(t)R_2(t) = D_2(t)R_1(t)D_3(t)$. Hence $A(t) < B(t)$, similarly $A(t) < C(t)$ and the lemma is proved.

We may now prove Theorem 2:

Proof of Theorem 2. By Lemma 3 total reductions $P_0(t)$ and $Q_0(t)$ can always be found. If $P_0(t) = S^{-1}Q_0(t)S$ for some permutation matrix S, then $P_0(t) < Q_0(t)$ and so P^t and Q^t are related. Suppose that for matrices $P_1(t)$ and $P_2(t)$ we have that $P_1(t) < P_2(t)$ and

$$P_2(t) = R(t)D(t) \quad \text{and} \quad P_1(t) = D(t)R(t).$$

The division matrix $D(0)$ defines a topological conjugacy between the one-sided subshifts of finite type $(\Sigma^+_{P^0_2}, \sigma_{P^0_2})$ and $(\Sigma^+_{P^0_1}, \sigma_{P^0_1})$ (cf. [3]). This topological conjugacy will preserve the measures given by $P_1(1)$ and $P_2(1)$ on $\Sigma^+_{P^0_1}$ and $\Sigma^+_{P^0_2}$ respectively (cf. [1]). By composing all the block-isomorphisms given by the division matrix, we conclude that there exists a block-isomorphism from $\Sigma^+_{P^0}$ onto $\Sigma^+_{Q^0}$.

Conversely, let ϕ be a block isomorphism from $\Sigma^+_{P^0}$ onto $\Sigma^+_{Q^0}$. Then by Lemma 2 and Lemma 3, $P_0(t)$ and $Q_0(t)$ are related by a string of matrices $P_0(t), P_1(t), \ldots, P_n(t) = Q_0(t)$. These can be thought of as vertices of a polygonal line (see Fig. 1) with a side joining $P_i(t)$ to $P_{i+1}(t)$ up to the right if $P_i(t) < P_{i+1}(t)$ and down to the right if $P_{i+1}(t) < P_i(t)$. If $P_i(t)$ and $P_{i+1}(t)$ are conjugate by a permutation matrix, then we draw a horizontal line.

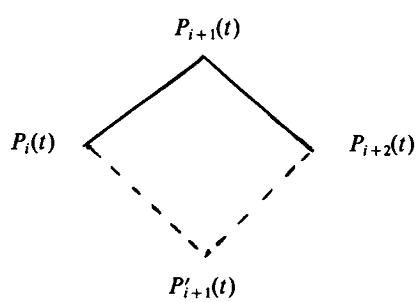

Fig. 1.

Using Lemma 4, any peak vertex of this graph can be lowered to obtain a lowest graph connecting $P_0(t)$ to $P_n(t) = Q_0(t)$. This lowest graph cannot contain a local minima for then there would be a strictly smaller total reduction of $P_0(t)$ and $Q_0(t)$. Hence $P_0(t)$ and $Q_0(t)$ are related by a permutation matrix.

References

1. W. Parry and S. Tuncel, *On the stochastic and topological structure of Markov-chains*, Bull. London Math. Soc. **14** (1982), 16–27.
2. W. Parry and R. F. Williams, *Block coding and a zeta function for finite Markov chains*, Proc. London Math. Soc. **35** (1977), 483–495.
3. R. F. Williams, *Classification of subshifts of finite type*, Ann. of Math. **98** (1973), 120–153; *Errata*, Ann. of Math. **99** (1974), 380–381.